

Open Research Online

The Open University's repository of research publications and other research outputs

Uncertainty handling in the context of ontology mapping for question answering

Conference or Workshop Item

How to cite:

Nagy, Miklos; Vargas-Vera, Maria and Motta, Enrico (2006). Uncertainty handling in the context of ontology mapping for question answering. In: AAAI-2006 Symposium on Semantic Web for Collaborative Knowledge Acquisition., 12-15 Oct 2006, Arlington, VA, USA.

For guidance on citations see [FAQs](#).

© 2006 The Authors

Version: Version of Record

Link(s) to article on publisher's website:

<http://www.cild.iastate.edu/events/aaai06symposium.html>

Copyright and Moral Rights for the articles on this site are retained by the individual authors and/or other copyright owners. For more information on Open Research Online's data [policy](#) on reuse of materials please consult the policies page.

oro.open.ac.uk

Uncertainty Handling in the Context of Ontology Mapping for Question-Answering

Miklos Nagy, Maria Vargas-Vera and Enrico Motta

Knowledge Media Institute (KMi)

The Open University

Walton Hall, Milton Keynes, MK7 6AA, United Kingdom

mn2336@student.open.ac.uk; {m.vargas-vera, e.motta}@open.ac.uk

Abstract

This paper describes a framework for integrating similarity measures and Dempster-Shafer belief functions for data integration in the context of multi agent ontology mapping. In order to incorporate uncertainty inherent to the ontology mapping process, we propose utilizing the Dempster-Shafer model for dealing with incomplete and uncertain information produced during the mapping. A novel approach is presented how assessing belief can influence the similarities originally created by both syntactic and semantic similarity algorithms. Our approach is an alternative to the classical Bayesian reasoning which has been investigated for improving the efficiency of creating ontology mappings.

Introduction

With the continuously increasing amount of data produced by electronic systems the integration of data and knowledge from multiple heterogeneous sources is an important problem that Scientific Database Community is facing today. To share information different multi agent architectures (Panti, Penserini, and Spalazzi 2002; Purvis, et al. 2003; Zhang, et al. 2004) have been proposed that utilize ontologies and ontology mapping as a source of the common knowledge. These architectures are the alternative to the federated approach, which is used to allow scientists to maintain control of their data, while sharing it within the community. An important aspect of ontology mapping in the context of knowledge management of heterogeneous scientific databases is how the incomplete and uncertain results of the different similarity algorithms can be interpreted during the mapping process. This has started to become a well-acknowledged research direction. As the latest research started moving towards a more automated mapping process it has been recognized that current approaches do not fully investigate the nature of the resultant similarity information and mainly rely on human domain expert to make a judgment about the correctness of the established mapping. However in the context of question answering like the AQUA (Vargas-Vera, Motta 2004; Vargas-Vera, Motta, and Domingue 2003) system the dynamic nature of the source information (e.g. web

enabled databases) does not make it possible to re-apply domain expert help every time the source changes. Additionally considering the dynamic nature of the Semantic web that is the extension of the current World Wide Web (WWW) it is hardly imaginable that isolated applications will be able to serve successfully the users' ever growing requirements since the information available to human decision makers continues to grow beyond human cognitive capabilities. In such an environment a single agent or application limited by its knowledge, perspective and its computational resources cannot cope with the before mentioned scenarios effectively. As the domain becomes larger and more complex, open and distributed, a set of cooperating agents are necessary in order to address the reasoning task effectively. In this context each agent carries only partial knowledge of the domain and can observe it from its own perspective where available prior knowledge is generally uncertain. Our main argument is that knowledge cannot be viewed as a simple conceptualization of the world, but it has to represent some degree of interpretation. Such interpretation depends on the context of the entities involved in the process. This idea is rooted in the fact the different entities' interpretations are always subjective, since they occur according to an individual schema, which is then communicated to other individuals by a particular language. Our novel approach utilizes a multi agent framework to address the before mentioned problems. Different mapping agents provide similarity measures about particular entities (e.g. material, specimen, etc.) and uncertainty plays a central role interpreting such similarities. Once we obtained different similarities we need to combine them in order to achieve more coherent and reliable results. To do this we use the Dempster-Shafer theory of evidence (Shafer 1976) that can be used to assess and combine the belief in the correctness of the different similarity measures based on subjective probabilities. Our approach also does not assume the existence of global or reference ontology that is the superset of the different source ontologies and contains the existing mappings a priori. The approach described here makes it possible to perform effective query answering with multiple source ontologies. Our first experimental system considers query answering over Web enabled S&T

(Scientific and Technical) or engineering databases which are described with their own domain specific ontologies. The paper is organized as follows:

Section ontology mapping in a multi agent system presents a short overview of the mapping framework and describes how mapping agents carry out the mapping process at different levels. Section similarity introduces the similarity algorithm used by the framework to assess syntactic and semantic similarities between the posed query and the local ontologies. Section uncertainty describes how the problem of uncertain information created by the similarity mapping process is resolved and handled by the mapping framework. Section evaluation of the system presents our experiments using the data from the Ontology Alignment Evaluation Initiative. Next section called strengths and limitations of the system sums up the advantages and disadvantages of our approach. Section related work discusses similar work and section conclusions provides our conclusions and describes future research directions.

Ontology Mapping in a Multi Agent System

In real scenario's ontology mapping can be carried out on domains with large number of classes and properties. To achieve the necessary performance for a real time mapping we utilize multi agent architecture. Without the multi agent architecture the response time of the system can increase exponentially when the number of concepts to map increases due to the Dempster's rule of combination. The high-level system architecture (figure 1) shows how the functional parts of the system are related with each other.

Our architecture consists of the following components:

1. Data: On the data layer the heterogeneous data sources are represented by their ontologies.
2. Mediator: In the mediator layer the agents are organized in different levels. Agents at the broker level are responsible for decomposing the query into sub queries, based on the meta-descriptors. The meta-descriptor is the key component of the system that describes what kind of information can be found in the different sources. This is practically FOL knowledge base that contains information about relations of the local resources. As an example let's consider a query where one is looking for materials with a specific name. The meta-descriptor will contain information that e.g. Ontology1 and Ontology2 describe Material related entities. The decomposed query parts are sent to the mapping agents in the mapping layer. Mapping agents obtain the relevant information from the sources through source agents. Agents communicate through the blackboard which is a task independent architecture for integrating multiple knowledge sources e.g. different local agents. Task independent means that it can be used for a wide range of tasks. In a blackboard system, a set of knowledge sources share a common

global database (blackboard). The contents of the blackboard are often called hypotheses. Knowledge sources respond to changes on the blackboard, and interrogate and subsequently directly modify the blackboard. This modification results in the creation, modification and solution of hypotheses. Because of only knowledge sources are allowed to make changes to the blackboard it is through the blackboard how the knowledge sources communicate and cooperate. The blackboard holds the state of the problem solution, while the knowledge sources make modifications to the blackboard when appropriate.

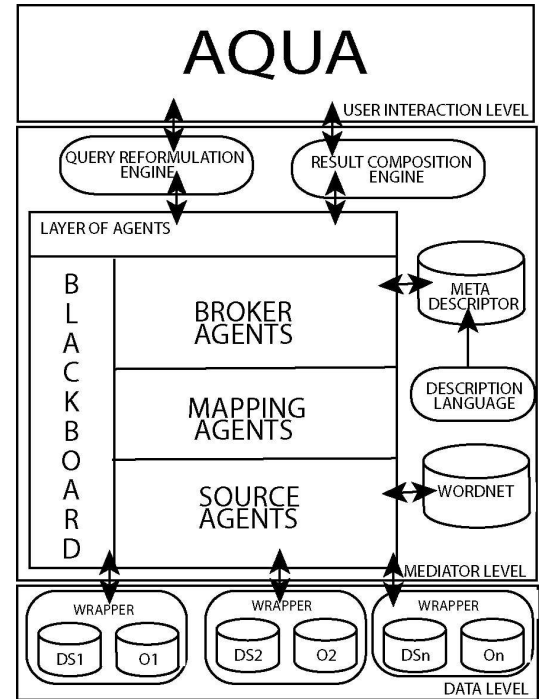


Figure 1. High level system architecture

3. User interaction: The AQUA query answering system itself, which provides precise answers to specific questions raised by the user. Users write a query in English, and then AQUA translates the query in First Order Logic (FOL) predicates. Consider the query: "What kind of tests has been carried out on material that has name 10 CrMo 9 10?" This would be represented in FOL as: $(\exists x, \forall y)(\text{test}(x) \text{ and material}(y) \text{ and hasName}(y, 10 \text{ CrMo } 9 \text{ 10}) \text{ and carriedOutOn}(x, y))$

Similarity

Syntactic Similarity

To assess syntactic similarity between ontology entities we use different string-based techniques to match names and name descriptions. These distance functions map a pair of

strings to a real number, which indicates a qualitative similarity between the strings. To achieve more reliable assessment we combine different string matching techniques such as edit distance like functions e.g. Monger-Elkan (Monge, and Elkan 1996) to the token-based distance functions e.g. Jaccard (Cohen, Ravikumar, and Fienberg 2003) similarity. To combine different similarity measures we use Dempster's rule of combination (see in section uncertainty). There are several reasonable similarity measures exist, each being appropriate to certain situations. To maximize our system's accuracy we employ a broad variety of similarity measures. At this stage of the similarity mapping our algorithm takes one entity from Ontology 1 and tries to find similar entity in extended query. The similarity mapping process is carried out on the following entities:

- Concept-name similarity
- Property set similarity

The use of string distances described here is the first step towards identifying matching entities between query and the ontology or between ontologies with little prior knowledge, or ill structured data. However, string similarity alone is not sufficient to capture the subtle differences between classes with similar names but different meanings. So we work with WordNet in order to exploit synonymy at the lexical-level. Once our query sting is extended with lexically synonym entities we calculate the string similarity measures between the query and the ontologies. In order to increase the correctness of our similarity measures the obtained similarity coefficients need to be combined. Establishing this combination method was our primary objective that had been included into the system. Further once the combined similarities have been calculated we developed a simple methodology to derive the belief mass function that is the fundamental property of Demster-Shafer evidence theory.

Semantic Similarity

For semantic similarity between concept, relations and the properties we use graph-based techniques. We take the extended query and the ontology input as labeled graphs. The semantic matching is viewed as graph-like structures containing terms and their inter-relationships. The similarity comparison between a pair of nodes from two ontologies is based on the analysis of their positions within the graphs. Our assumption is that if two nodes from two ontologies are similar, their neighbours might also be somehow similar. We consider semantic similarity between nodes of the graphs based on similarity of leaf nodes. That is, two non-leaf schema elements are semantically similar if their leaf sets are highly similar, even if their immediate children are not.

The main reason why semantic heterogeneity occurs in the different ontology structures is because different institutions develop their data sets individually, which as a result contain many overlapping concepts. Assessing the

above-mentioned similarities in our multi agent framework we adapted and extended the SimilarityBase and SimilarityTop algorithms (Vargas-Vera, and Motta 2004; Vargas-Vera, and Motta 2004) used in the current AQUA system for multiple ontologies. Our aim is that the specialized agents will simulate the way a human designer would describe a domain based on a well-established dictionary. What also needs to be considered when the two graph structures are obtained from both the user query fragment and the representation of the subset of the source ontology is that there can be a generalization or specialization of a specific concepts present in the graph which was obtained from the local source and this needs to be handled correctly. In our multi agent framework the extended and combined SimilarityBase and SimilarityTop algorithms can be described as follows:

1. Using WordNet an extended directed graph is constructed from the FOL query fragment where there are bi-directional edges between the nodes representing the concepts and there are directed edges from the concepts to the property nodes. In this step the specialized agents try to determine all possible meanings of the query fragment. Figure 2 depicts the graph representation of the material(y) and hasName(material, 10 CrMo 9 10) FOL query fragment.

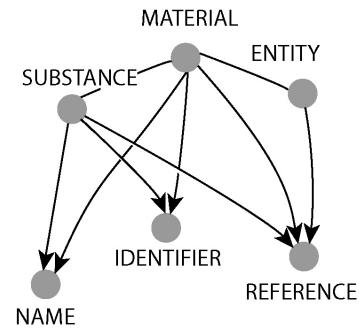


Figure 2. G0 query fragment graph

2. Based on different string similarity measures (see in section syntactic similarity) the specialized agent builds up a directed graph from the local ontology structures that can be an answer the particular query fragment. We consider the similarity measures as subjective probabilities and therefore to calculate the belief mass function we normalize the state space to 1 for the particular query fragment as required by the evidence theory (see belief mass function in section uncertainty). Then the belief function can be calculated for each concept and property. Figure 3 depicts two graph fragments with belief functions obtained from two different ontologies where G1 is the graph located to left hand side of the middle bar

and G2 is located on the right hand side. G1 and G2 graph fragments have been extracted from the different local ontologies and have been identified as possible answers that correspond to the G0 query.

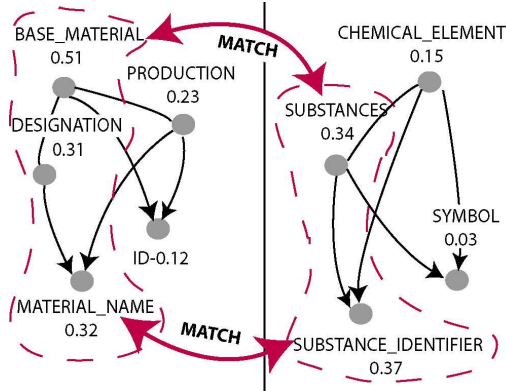


Figure 3. G1 and G2 graph belief function representations of the local ontology fragment.

- Top-down sub-graph (isomorphism) similarity assessment is applied on the graph G0 in order to find the sub graph G1 and G2 with the highest sum of belief functions respectively. The aim is to find identical sub graphs with the highest sum of belief functions like G1 and G2 in order to assess the similarity of the concepts and properties that can answer the query fragment. The search for the sub graphs starts from the concept nodes towards property nodes through the directed edges. Once we reached the property node the search stops. If the path with the highest sum of belief has produced a sub graph identical (isomorph) to G1 and G2 than the agent can deduce that the query fragment can be answered from the sources that belong to the particular ontology and the concepts or properties identified in the different sources are similar to both each other and to the query fragment. Belief functions express the extent of belief in the existence of the similarity mappings. In case G1 or G2 contains nodes that could not be found in the G0, because of the nature of the top down assessment the agent can deduce that the particular concept node is a specialization of the concept that was identified by the algorithm.

Uncertainty

In our framework we use the Dempster-Shafer theory of evidence, which provides a mechanism for modeling and reasoning uncertain information in a numerical way particularly when it is not possible to assign a belief to a single element of a set of values. Consequently the theory allows the user to represent uncertainty for knowledge representation, because the interval between support and

plausibility can be easily assessed for a set of hypotheses. Missing data also could be modeled by Dempster-Shafer approach and additionally evidences from two or more sources can be combined using Dempster's rule of combination. The combined support, plausibility, disbelief, and uncertainty can each be separately evaluated. The main advantage of the Dempster-Shafer theory over the classical probabilistic theories is the evidence of different levels of abstraction can be represented in a way, which allows clear distinction to be made between uncertainty and ignorance. Further advantage is that the theory provides a method for combining the effect of different learned evidences to establish a new belief by using Dempster's combination rule. Let's first describe the basic concepts of the Dempster-Shafer theory and how it corresponds to our system.

Frame of Discernment (Θ): A finite set representing the space of hypotheses. It contains all possible mutually exclusive context events of the same kind. In our system this corresponds to the possible properties, those of the base entities that describe the domain e.g. Material Name, Test Control, and Specimen Identifier etc.

Evidence: available certain fact and is usually a result of observation. Used during the reasoning process to choose the best hypothesis in Θ . In our system this can be a certain observation e.g. in the case of a material: the production details have been observed or not.

Belief mass function (m): is a finite amount of support assigned to the subset of Θ . It represents the strength of some evidence and

$$\sum_{A \subseteq \Theta} m(A) = 1$$

where $m(A)$ is our exact belief in a proposition represented by A . The similarity algorithms itself produce these assignment based on the above mentioned (see in section similarity) similarities e.g. between name and identifier property the assigned value is 0.7.

Once the belief mass functions have been assigned the following additional measures can be derived from the available information.

Belief: amount of justified support to A that is the lower probability function of Dempster, which accounts for all evidence E_k that supports the given proposition A .

$$belief_i(A) = \sum_{E_k \subseteq A} m_i(E_k)$$

Plausibility: amount of potential support for "A" that is the upper probability function of Dempster, which accounts for all the observations that do not rule out the given proposition.

$$plausibility_i(A) = 1 - \sum_{E_k \cap A = \emptyset} m_i(E_k)$$

Ignorance: the lack of information.

$$ignorance(A) = plausibility(A) - belief(A)$$

Once all the necessary variables have been assigned to a qualitative value we need to combine the belief mass functions that were created by the different agents for the particular query fragment.

Dempster's rule of combination:

Suppose we have two mass functions $m_i(E_k)$ and $m_j(E_k')$ and we want to combine them into a global $m_{ij}(A)$. Following Dempster's combination rule

$$m_{ij}(A) = m_i \oplus m_j = \sum_{E_k \cap E_k'} m_i(E_k) * m_j(E_k')$$

An important aspect of the mapping is how one can make a decision over how different similarity measures can be combined and which nodes should be retained as best possible candidates for the match. To combine the qualitative similarity measures that have been converted into belief mass functions we use the Dempster's rule of combination and we retain the node which belief function has the highest value. This process called belief revision which is rearranging a cognitive state in order to embody new information while preserving the global consistency. This process is computationally very expensive and from an engineering point of view, this means that it not always convenient or possible to build systems in which the belief revision process is performed globally by a single unit. Therefore, applying multi agent architecture is an alternative and distributed approach to the single one, where the belief revision process is no longer assigned to a single agent but to a group of agents, in which each single agent is able to perform belief revision and communicate with the others. The main difference is that in this case there is no more a single agent having a global view of the system, but each agent has partial view of it. This allows that the computational load to be divided among the agents of the group. Our algorithm takes all the concepts and its properties from the different ontologies and assesses similarity with all the concepts and properties in the query graph. Imagine the scenario mentioned in the semantic similarity section. In the ontology graph G1 we take the node "BASE_MATERIAL" and utilizing different similarity measures (see in section syntactic similarity) we receive two similarity graphs where figure 4 depicts the CharJaccard and the Monger-Elkan similarity measures that has been obtained by comparing the "BASE_MATERIAL" node to concepts in our extended query graph. CharJaccard Similarity uses letter sets from the comparison instances to evaluate similarity and the Monge Elkan approach makes other additional tests by taking the semantic similarity of a number of fields and sub fields into consideration. The above-mentioned similarities are calculated as follows:

$$CharJaccard(S, T) = \frac{|S \cap T|}{|S \cup T|}$$

and

$$Monge - Elkan(S, T) = \frac{1}{|S|} \sum_{i=1}^{|S|} \max_{j=1}^{|T|} Monge - Elkan(S_i, T_j)$$

where S and T represents different terms such as concepts or properties that needs to be compared.

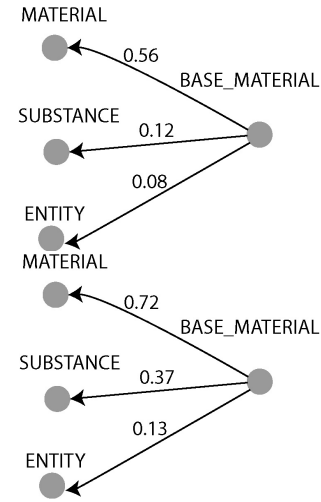


Figure 4. Obtained similarities based on CharJaccard(top) and Monger-Elkan(down) similarity

To obtain more reliable results we need to combine the similarity assessments that have been produced by the different similarity algorithms (figure 4). The combination of the different similarity measures is not a straightforward question and it includes several biases. Our approach is to consider these measures as subjective probabilities and utilize a well-established framework that provides convenient way to represent and combine our evidences. Particularly appealing in Dempster-Shafer theory is the ability to model ignorance as well as uncertainty in the particular system. However because n belief formula lead to a mass function of 2^n combinations, keeping a mass function in memory and updating it when the belief base changes, will lead to a combinatorial explosion in both processing time and memory requirements and this is a serious limitation of any practical applications of this theory. To overcome these limitations we need to utilize multi agent architecture so the problem space can be distributed between the agents thus making computation viable for domains with large number of variables. We believe that Dempster-Shafer theory is a convenient and intuitive way to model uncertainty of agent beliefs the computational complexity can be controlled.

The Dempster's rule of combination provides a well-established approach however it works with belief mass functions where the sums of the masses add up to 1. To convert similarities into belief masses we need to normalize the problem space into 1. This way we can easily obtain the necessary masses so we can utilize the combination rule.

Instead of just retaining nodes where the belief mass function exceeds a certain predefined limit we can examine leaf nodes from the graph and calculate a belief that will give us a good indication which node needs to be retained as the best possible matching candidate.

Using the above-mentioned process we can generate the combination of the similarity measures in the graphs (figure 4) with the necessary belief mass functions which can be used to calculate the belief in a particular proposition e.g. the highest belief shows that MATERIAL and BASE_MATERIAL is a definite match. It is important to note that we have illustrated our example using material ontologies for sake of simplicity. However, these ontologies are not part of the Alignment Evaluation Initiative. Therefore, these evaluations are not presented in Figure 5.

Evaluation of the System

Experiments

We have performed computer experiments with the existing benchmarks of the Ontology Alignment Evaluation Initiative (Euzenat, Stuckenschmidt, and Yatskevich 2005) which is an international initiative that has been set up for evaluating ontology matching algorithms. We have not participated in the contest but used the provided ontologies in the systematic benchmarks and compared our mapping results to the contestant's results.

Our main objective was to compare our system and algorithms to existing approaches on the same basis and to allow drawing constructive conclusions. The benchmarks contain tests which were systematically generated to as to start from some reference ontology and discarding a number of information in order to evaluate how the algorithm behave when this information is lacking. The reference ontology contained 33 named classes, 24 object properties, 40 data properties. Further each generated ontology was aligned with the reference ontology. The evaluation was measured with recall and precision which are useful measures that have a fixed range and are easy to compare across queries and engines.

By definition precision is a measure of the usefulness of a hitlist and calculated as

$$precision = \frac{\text{number of relevant hits in hitlist}}{\text{number of hits in hitlist}}$$

where hitlist is an ordered list of hits in decreasing order of relevance to the query.

and recall is a measure of the completeness of the hitlist

$$recall = \frac{\text{number of relevant hits in hitlist}}{\text{number of relevant document in collection}}$$

Recall is a measure of how well the engine performs in finding relevant entities. Recall is 100% when every relevant entity is retrieved. However it is possible to achieve 100% by simply returning every entity in the collection for every query. Therefore, recall by itself is not a good measure of the quality of a search engine. Precision is a measure of how well the engine performs in not

returning non relevant documents. Precision is 100% when every entities returned to the user is relevant to the query. There is no easy way to achieve 100% precision other than in the trivial case where no document is ever returned for any query. Both precision and recall have a fixed range: 0.0 to 1.0 (or 0% to 100%). A good mapping algorithm must have a high recall to be acceptable for most applications. The most important factor in building better mapping algorithms is to increase precision without worsening the recall. In our evaluation we have tested our similarity algorithm called "DSSim" with the following ontologies (test) from the benchmark:

- Nr. 103 - Concept test: Language generalization. This test compares the ontology with its generalisation in OWL Lite (i.e., unavailable constraints are replaced by the more general available). The generalization basically removes owl:unionOf and owl:oneOf and the Property types (owl:TransitiveProperty).
- Nr. 204 - Systematic: Naming conventions. Different naming conventions (Uppercasing, underscore, dash, etc.) are used for labels. Comments have been suppressed.
- Nr. 205 - Systematic: Synonyms. Labels are replaced by synonyms. Comments have been suppressed.
- Nr. 221 - Systematic: No hierarchy. All subclass assertions to named classes are suppressed.
- Nr. 222 - Systematic: Flattened hierarchy. A hierarchy still exists but has been strictly reduced.
- Nr. 223 - Systematic: Expanded hierarchy. Numerous intermediate classes are introduced within the hierarchy.
- Nr. 301 - Real ontology: BibTeX(MIT). The BibTeX is the starting point for a useful bibliographic ontology. This is a test of comparing our test ontology with an actual ontology, simpler and closer to the initial BibTeX ontology. The alignment result contains some inclusion (<) alignment relations.

algo	DSSim		falcon		foam		ctxMatch2-		cms	
test	Prec.	Rec.	Prec.	Rec.	Prec.	Rec.	Prec.	Rec.	Prec.	Rec.
103	0.96	1	1	1	0.98	0.98	0.87	0.34	0.67	0.25
203	0.96	1	1	1	1	1	0.87	0.34	1	0.24
204	0.62	0.65	1	1	1	0.97	0.87	0.28	1	0.24
205	0.42	0.43	0.88	0.87	0.89	0.73	0.36	0.04	1	0.09
221	0.96	1	1	1	1	1	0.87	0.34	1	0.27
222	0.91	1	1	1	0.98	0.98	0.73	0.26	1	0.23
223	0.95	1	1	1	0.99	0.98	0.83	0.31	0.96	0.26
301	0.23	0.38	0.96	0.8	0.83	0.31	0	0	1	0.13

Figure 5. Comparison of different similarity algorithms

Figure 5 depicts the comparisons where the similarity algorithm abbreviations stand for the followings:

- Falcon (Jian, et al. 2005) Falcon-AO is an automatic ontology matching tool
- Foam (Ehrig, and Sure 2005) Framework for Ontology Alignment and Mapping

- ctxMatch (Bouquet, et al. 2003) ontology mapping tool from University of Trento
- cms (Kalfoglou, and Hu 2005) CROSI (Ontology) Mapping System (CMS)

For the complete list of alignment results one can refer to Ontology Alignment Initiative web-site¹. We have implemented our similarity mapping algorithm called "DSsim" and embedded into Ontology Alignment Framework which provide the complete test base, including couples of ontologies to align as well as expected results. Our similarity algorithm implemented the Ontology Alignment API interfaces which enable the integration of the particular similarity mapping and evaluation algorithms. The API is used to create an alignment object, providing the two ontologies, calling the align method(in this case DSsim). The alignment object then bears the result of the alignment procedure. Experiments show, that our algorithm performs well comparing to another systems especially when syntactic differences are dominant in the ontology. For the semantic differences where e.g. synonyms were used (test nr. 205) or with the real word ontology (test nr. 301) our approach underperformed comparing to the "Foam" or "Falcon" algorithms. The explanation for this can be summarized as follows:

- DSsim algorithm uses the general WordNet dictionary for these particular benchmarks which does not contain extensive synonym set for the concepts or properties that are in the test ontologies.
- Because of the computational complexity of the combination algorithm the possible variable number i.e. the problem space is limited to a certain size which negatively influences the precision of the similarity calculation.

Strengths and Limitation of the System

Our mapping algorithm takes mainly concepts (classes) and properties (object and data type) to capture the specific restrictions in the particular ontologies and converts them into directed graph fragments. As a consequence our method is not heavily dependent on subclass, sub property, disjointness or equivalency relationships among classes and properties hence on the logical constraints imposed by the ontology language itself. Additionally the query terms are extended with their synonyms from WordNet so the uncertainty can be distributed sufficiently that can lead to a large number of possibly valid choices. However since Dempster's combination rule is computationally really expensive operation we need to reduce the problem space therefore the number of additional variables per query fragment. This can lead to the loss of valuable information and consequently more irrelevant mappings.

Related Work

Ontology mapping is widely investigated area and numerous approaches have led to different solutions.

Derived from the data engineering community several solutions have been proposed based on mediator architecture where logical database schemas are used as shared mediated views over the queried schemas. A number of systems have been proposed e.g. TSIMMIS (Garcia-Molina, et.al. 1997), Information Manifold (Halevy 1998), InfoSleuth (Bayardo, et al. 1997), MOMIS (Beneventano, Bergamaschi, Guerra, and Vincini 2001) that shows the flexibility and the scalability of these approaches. Derived from the knowledge engineering community the use of ontologies (conceptual domain knowledge schemas) is the main approach for resolving semantic differences in heterogeneous data sources.

Until recently, the focus of the researchers was not on uncertainty handling during the mapping since initially only different logic (FOL, Description Logics) based approaches had been utilized. As practical application of ontologies emerged on the web it has been acknowledged that because of the dynamic nature of the Web the problem of inconsistencies, controversies and lack of information needs to be handled. First systems that used probabilistic information like LSD, GLUE (Doan, Madhavan, Domingos, and Halevy 2002) proved that combining different similarity measures based on their probability could significantly improve the accuracy of the mapping process. It is worth to note that the Bayesian networks and different variants dominate current research addressing the qualitative reasoning and decision-making problem under uncertainty. Although these approaches produced numerous successful real world applications there are several situations where the problem cannot be represented properly within the classical probability framework. The most related research direction for ontology mapping framework under uncertainty using Bayesian networks (Zhongli, Yun, and Rong 2004) to tackle this problem. In the last several years, a numerous research papers (Schoken, and Hummel 1993; Teixeira, and Milidui 1993; Lalmas 1997) have been published on Dempster-Shafer theory related to traditional information retrieval where document is considered as an atomic entity that is indexed and retrieved as a whole by the system, and is presented to the user as a query result.

Conclusions

The increasing popularity of the Semantic Web poses new challenges for ontology mapping. If we accept that mapping ontologies can provide a better knowledge management of the heterogeneous sources on the Semantic Web, then issues of inconsistency and incompleteness need to be addressed. The above mentioned issues are generic enough and affect the Semantic Web as a whole, hence

¹ <http://oaei.ontologymatching.org/2005/results/>

ontology mapping systems that operate in this environment should have the appropriate mechanisms to cope with these issues. In this complex environment different scientific disciplines need to be utilized together to achieve better results for answering user queries within an acceptable response times. We think that in our implementation we have made an encouraging step towards a theoretical solution but the different key system components such as similarity measure or the scalability of uncertainty handling part needs to be investigated further. We believe that probability theory and distribution does not have enough expressive power to tackle certain aspects of the uncertainty e.g. total ignorance. The main contribution of this paper and our research is the application of Dempster-Shafer theory for the ontology mapping problem. Therefore we expect that evidence theory is the most suitable approach and needs to be investigated in ontology mapping context though this has not been done so far. We believe this is because Dempster-Shafer combination rules can be unfeasible in domains with large number of variables. Different optimisation methods have been developed but to date we could not find similar approaches that considered multi agent environments. Local computation and valuation networks use joint tree structures to narrow down the number of focal elements and different architectures has been proposed based on message passing schemas to carry our inference and resolve the problem of the Dempster's rule of combination. In our future research we will investigate how these optimisation methods can be adapted and applied in our scenario with a dynamic multi agent environment where each agent has partial knowledge of the domain.

References

- Bayardo, R.; et al. 1997. Infosleuth, Agent-based Semantic Integration of Information in Open and Dynamic Environments. In *Proceedings of ACM SIGMOD Conference on Management of Data*, Tucson, Arizona.
- Beneventano, D.; Bergamaschi, S.; Guerra, F.; Vincini, M. 2001. The MOMIS Approach to Information Integration. In *Proceedings of 3rd International Conference on Enterprise Information Systems*, Setubal, Portugal.
- Bouquet, P.; et al 2003. A SAT-Based Algorithm for Context Matching. Technical Report DIT-03-005, Informatica e Telecomunicazioni, University of Trento.
- Cohen, W.; Ravikumar, P.; and Fienberg, S. 2003. A Comparison of String Distance Metrics for Name-Matching Tasks, In *Proceedings of Information Integration on the Web (IIWeb 2003)*, Accapulco, Mexico.
- Doan, A. H.; Madhavan, J.; Domingos, P.; Halevy A. 2002. Learning to Map between Ontologies on the Semantic Web. In *Proceedings of Eleventh International World Wide Web Conference (WWW 2002)*, Hawaii, USA.
- Ehrig, M.; and Sure, Y. 2005. FOAM - Framework for Ontology Alignment and Mapping - Results of the Ontology Alignment Evaluation Initiative, In *Proceedings of the K-CAP 2005 Workshop on Integrating Ontologies*, Banff, Canada.
- Euzenat, J.; Stuckenschmidt, H.; and Yatskevich, M. 2005. Introduction to the ontology alignment evaluation 2005. In *Proceedings of K-CAP 2005 Integrating Ontologies Workshop*, Banff, Canada.
- Garcia-Molina, H.; et.al. 1997. The TSIMMIS Approach to Mediation: Data Models and Languages, *Journal of IEEE Intelligent Systems*, 8(2):117-132.
- Halevy, A. 1998 The Information manifold approach to data integration, *Journal of IEEE Intelligent Systems*, 1312-16.
- Jian, N.; et al. 2005. FalconAO: Aligning Ontologies with Falcon, In *Proceedings of the K-CAP 2005 Workshop on Integrating Ontologies*, Banff, Canada.
- Kalfoglou, Y.; and Hu, B. 2005. CROSI Mapping System (CMS) - Result of the 2005 Ontology Alignment Contest.
- Lalmas, M. 1997. Dempster-Shafer's Theory of Evidence Applied to Structured Documents: Modelling Uncertainty. In *Proceedings of SIGIR*, ACM, New York.
- Monge, A. E.; Elkan, C. P. 1996. The field-matching problem: algorithm and applications. In *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining*, Portland, US.
- Panti, M.; Penserini, L.; and Spalazzi, L. 2002. A pure P2P approach to information integration, Tec. Report, Computer Science dep., Polytechnic University of Marche, Italy.
- Purvis, M.; et al. 2003. A multi-agent system for the integration of distributed environmental information. *Environmental Modelling & Software*, Volume 18, Issue 6.
- Schoken, S. S.; and Hummel, R. A. 1993. On the use of the Dempster-Shafer model in information indexing and retrieval applications. *Int. Journal of Man-Machine Stud.*, 39, 1-37.
- Shafer, G. 1976. A Mathematical Theory of Evidence.: Princeton University Press.
- Teixeira, de Silva W.; and Milidui, R. L. 1993. Belief function model for information retrieval. *Journal of American Society of Information Science*, 44, 10-18.
- Vargas-Vera, M. and Motta, E. 2004. AQUA - Ontology-based Question Answering System. In *Proceedings of Third International Mexican Conference on Artificial Intelligence (MICAI-2004)*, Mexico city, Mexico.
- Vargas-Vera, M.; Motta, E.; and Domingue, J. 2003. AQUA: An Ontology-Driven Question Answering System. AAAI Spring Symposium, New Directions in Question Answering, Stanford, USA.: AAAI Press.
- Vargas-Vera, M.; and Motta, E. 2004. A Knowledge-Based Approach to Ontologies Data Integration. KMi-TR-152, The Open University, UK.
- Vargas-Vera, M.; and Motta, E. 2004. An Ontology-driven Similarity Algorithm. KMi-TR-151, Knowledge Media Institute, The Open University, UK.
- Zhang, Z; et al. 2004. Multiagent system solutions for distributed computing, communications, and data integration needs in the power industry, In *Proceedings of the General Meeting of the IEEE Power Engineering Society*, pp. 45-49.: IEEE Press.
- Zhongli, D.; Yun, P.; Rong, P. 2004. A Bayesian Approach to Uncertainty Modeling in OWL Ontology, In *Proceedings of International Conference on Advances in Intelligent Systems - Theory and Applications (AISTA2004)*. Luxembourg-Kirchberg, Luxembourg.